(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 August 2001 (30.08.2001)

PCT

(10) International Publication Number
WO 01/62893 A2

(51) International Patent Classification⁷: C12N

(21) International Application Number: PCT/US01/05992

(22) International Filing Date: 26 February 2001 (26.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/185,037 25 February 2000 (25.02.2000) US
60/223,828 8 August 2000 (08.08.2000) US

(71) Applicant *(for all designated States except US)*: **COREXA CORPORATION** [US/US]; 1124 Columbia Street, Suite 200, Seattle, WA 98104 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CAMPOS-NETO, Antonio** [BR/US]; 9388 N.E. Midship Court, Bainbridge Island, WA 98110 (US). **SKEIKY, Yasir** [CA/US]; 8327 25th Avenue, N.W., Seattle, WA 98117 (US). **OVENDALE, Pamela** [US/US]; 12021 30th Avenue S.E., Apt. 35, Everett, WA 98208 (US). **JEN, Shyian** [US/US]; 2345, 1/2 Boylston Avenue E 205, Seattle, WA 98122 (US). **LODES, Michael** [US/US]; 9223 36th Avenue S.W., Seattle, WA 98126 (US).

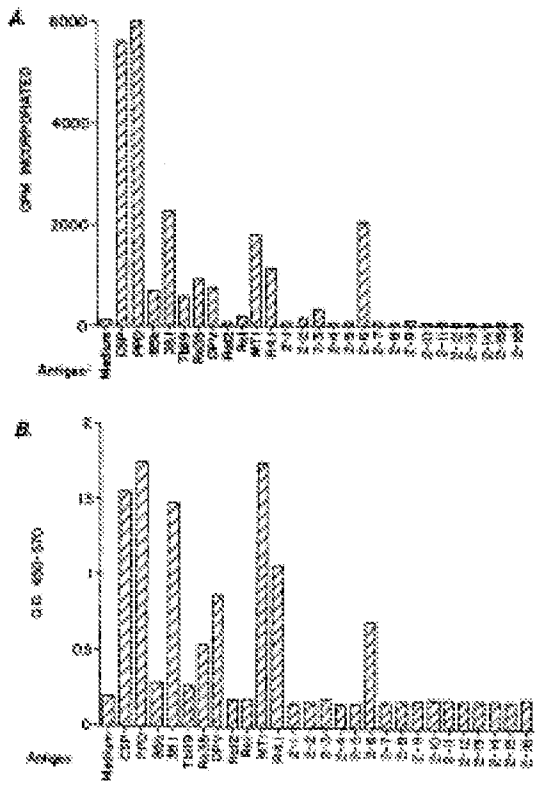
(74) Agents: **PARENT, Annette, S.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th floor, San Francisco, CA 94111-3834 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

(Continued on next page)

(54) Title: COMPOUNDS AND METHODS FOR DIAGNOSIS AND IMMUNOTHERAPY OF TUBERCULOSIS



(57) Abstract: Compounds and methods for diagnosing tuberculosis or for inducing protective immunity against tuberculosis are disclosed. The compounds provided include polypeptides that contain at least one immunogenic portion of one or more *Mycobacterium* proteins and DNA molecules encoding such polypeptides. Diagnostic kits containing such polypeptides or DNA sequences and a suitable detection reagent may be used for the detection of *Mycobacterium* infection in patients and biological samples. Antibodies directed against such polypeptides are also provided. In addition, such compounds may be formulated into vaccines and/or pharmaceutical compositions for immunization against *Mycobacterium* infection.

WO 01/62893 A2

patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *without international search report and to be republished upon receipt of that report*

COMPOUNDS AND METHODS FOR DIAGNOSIS AND IMMUNOTHERAPY OF TUBERCULOSIS

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application claims priority to patent application No. 60/185,037, filed 02/25/00; and patent application No. 60/223,828, filed August 8, 2000, herein each incorporated by reference in its entirety.

The present application is related to U.S. patent application Nos. 08/859,381, filed May 20, 1997 (abandoned); 08/858,998, filed May 20, 1997 (abandoned); 09/073,010, filed May 5, 1998; and 09/073,009, filed May 5, 1998; and to PCT application Nos. PCT/US98/10407, filed May 20, 1998; and PCT/US98/10514, filed May 20, 1998, herein each incorporated by reference in its entirety.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

Not applicable.

BACKGROUND OF THE INVENTION

Tuberculosis is a chronic, infectious disease, that is generally caused by infection with *Mycobacterium tuberculosis*. It is a major disease in developing countries, as well as an increasing problem in developed areas of the world, with about 8 million new cases and 3 million deaths each year. Although the infection may be asymptomatic for a considerable period of time, the disease is most commonly manifested as an acute inflammation of the lungs, resulting in fever and a nonproductive cough. If left untreated, serious complications and death typically result.

Although tuberculosis can generally be controlled using extended antibiotic therapy, such treatment is not sufficient to prevent the spread of the disease. Infected individuals may be asymptomatic, but contagious, for some time. In addition, although compliance with the treatment regimen is critical, patient behavior is difficult to monitor. Some patients do not complete the course of treatment, which can lead to ineffective treatment and the development of drug resistance.

Inhibiting the spread of tuberculosis will require effective vaccination and accurate, early diagnosis of the disease. Currently, vaccination with live bacteria is the most efficient method for inducing protective immunity. The most common Mycobacterium employed for this purpose is Bacillus Calmette-Guerin (BCG), an avirulent strain of *Mycobacterium bovis*. However, the safety and efficacy of BCG is a source of controversy and some countries, such as the United States, do not vaccinate the general public. Diagnosis is commonly achieved using a skin test, which involves intradermal exposure to tuberculin PPD (protein-purified derivative). Antigen-specific T cell responses result in measurable induration at the injection site by 48-72 hours after injection, which indicates exposure to Mycobacterial antigens. Sensitivity and specificity have, however, been a problem with this test, and individuals vaccinated with BCG cannot be distinguished from infected individuals.

While macrophages have been shown to act as the principal effectors of *M. tuberculosis* immunity, T cells are the predominant inducers of such immunity. The essential role of T cells in protection against *M. tuberculosis* infection is illustrated by the frequent occurrence of *M. tuberculosis* in AIDS patients, due to the depletion of CD4 T cells associated with human immunodeficiency virus (HIV) infection. Mycobacterium-reactive CD4 T cells have been shown to be potent producers of gamma-interferon (IFN-γ), which, in turn, has been shown to trigger the anti-mycobacterial effects of macrophages in mice. While the role of IFN-γ in humans is less clear, studies have shown that 1,25-dihydroxy-vitamin D3, either alone or in combination with IFN-γ or tumor necrosis factor-alpha, activates human macrophages to inhibit *M. tuberculosis* infection. Furthermore, it is known that IFN-γ stimulates human macrophages to make 1,25-dihydroxy-vitamin D3. Similarly, IL-12 has been shown to play a role in stimulating resistance to *M. tuberculosis* infection. For a review of the immunology of *M. tuberculosis* infection, *see* Chan and Kaufmann, in *Tuberculosis: Pathogenesis, Protection and Control*, Bloom (ed.), ASM Press. Washington, DC (1994).

Accordingly, there is a need in the art for improved diagnostic methods for detecting tuberculosis, as well as for vaccines and methods for preventing the infection. The present invention fulfills this need and further provides other related advantages.

SUMMARY OF THE INVENTION

Briefly stated, the present invention provides compounds and methods for preventing and diagnosing tuberculosis.

In one embodiment, polypeptides are provided that comprise an immunogenic portion of a *Mycobacterium* antigen, preferably a *Mycobacterium tuberculosis* antigen, or a variant of such an antigen that differs only in conservative substitutions and/or modifications, wherein the antigen comprises an amino acid sequence encoded by a polynucleotide having the nucleotide sequence recited in SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, the complements of said sequences, or a nucleotide sequence that hybridizes to the sequence set forth in SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, or an immunogenic fragment thereof. In a second embodiment, the present invention provides polypeptides comprising an immunogenic portion of a *Mycobacterium* antigen, preferably a *Mycobacterium tuberculosis* antigen, having the amino acid sequence described in SEQ ID NO:146, 161, or 163 or variants or immunogenic fragments thereof.

In related aspects, nucleotide sequences encoding the above polypeptides, recombinant expression vectors comprising these nucleotide sequences and host cells transformed or transfected with such expression vectors are also provided. In particular, the present invention provides an isolated polynucleotide that specifically hybridizes under moderately stringent conditions to a second polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164. In some embodiments, the isolated polynucleotide specifically hybridizes to the second polynucleotide under highly stringent conditions.

In another aspect, the present invention provides fusion proteins comprising a first polypeptide encoded by a polynucleotide having the sequence set forth in SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, or a fragment thereof, and a second polypeptide. In one embodiment, the first and second polypeptides are heterologous. Alternatively, the fusion proteins of the invention may comprise a first polypeptide encoded by a polynucleotide having a sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156,

157, 158, 159, 160, 162, and 164, or an immunogenic fragment thereof, and a known *Mycobacterium* antigen, preferably a *M. tuberculosis* antigen.

In further aspects of the subject invention, methods and diagnostic kits are provided for detecting *Mycobacterium* infection in a patient. The methods comprise contacting a biological sample with at least one of the above polypeptides and detecting in the sample the presence of antibodies that bind to the polypeptide or polypeptides, thereby detecting *Mycobacterium* infection in the biological sample. In a preferred embodiment, the *Mycobacterium* infection is a *M. tuberculosis* infection.

Suitable biological samples include whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid and urine. The diagnostic kits comprise one or more of the above polypeptides in combination with a detection reagent.

The present invention also provides methods for detecting *Mycobacterium* infection, comprising obtaining a biological sample from a patient, contacting the sample with at least one oligonucleotide primer in a polymerase chain reaction, the oligonucleotide primer being specific for a nucleotide sequence encoding the above polypeptides, and detecting in the sample a nucleotide sequence that amplifies in the presence of the first and second oligonucleotide primers. In one embodiment, the oligonucleotide primer comprises at least about 10 contiguous nucleotides of such a nucleotide sequence. In a preferred embodiment, the *Mycobacterium* infection is a *M. tuberculosis* infection.

In a further aspect, the present invention provides a method for detecting *Mycobacterium* infection in a patient, comprising obtaining a biological sample from the patient, contacting the sample with an oligonucleotide probe specific for a nucleotide sequence encoding the above polypeptides, and detecting in the sample a nucleotide sequence that hybridizes to the oligonucleotide probe. In one embodiment, the oligonucleotide probe comprises at least about 15 contiguous nucleotides of such a nucleotide sequence. In a preferred embodiment, the *Mycobacterium* infection is a *M. tuberculosis* infection.

In yet another aspect, methods are provided for detecting *Mycobacterium* infection in a patient, such methods comprising the steps of contacting a biological sample with a polypeptide, wherein the polypeptide comprises an amino acid sequence encoded by a polynucleotide having a nucleotide sequence selected from the group

consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, the complements of said sequences, or a nucleotide sequence that hybridizes to a sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, or an immunogenic fragment thereof, and detecting in the sample the presence of antibodies that bind to the polypeptide, thereby detecting *Mycobacterium* infection in the biological sample. In a preferred embodiment, the *Mycobacterium* infection is a *M. tuberculosis* infection. Diagnostic kits for use in such methods are also provided.

In another aspect, the present invention provides antibodies, both polyclonal and monoclonal, that bind to the polypeptides described above, as well as methods for their use in the detection of *Mycobacterium* infection.

Within other aspects, the present invention provides pharmaceutical compositions that comprise one or more of the above polypeptides, or a polynucleotide encoding such polypeptides, and a physiologically acceptable carrier or an adjuvant, *e.g.*, SBAS-2, QS-21, ENHANZYN (Detox), MPL, 3D-MPL, CWS, GM-CSF, SAF, ISCOMS, MF-59, RC-529, AS2, AS2', AS2'', AS4, AS6, TDM, AGP, CPG, Leif, saponin, and saponin mimetics, and derivatives thereof or mixtures thereof. In another aspect, the present invention provides pharmaceutical compositions that comprise one or more of the above polypeptides, or a polynucleotide encoding such polypeptides, and an adjuvant such as BCG. In another aspect the present invention provides methods in which one or more of the above polypeptides, or a polynucleotide encoding such polypeptides is administered to a subject who has been exposed to BCG. The invention also provides vaccines comprising one or more of the polypeptides as described above and a non-specific immune response enhancer, together with vaccines comprising one or more polynucleotides encoding such polypeptides and a non-specific immune response enhancer.

In yet another aspect, methods are provided for inducing protective immunity in a patient, comprising administering to a patient an effective amount of one or more of the above polypeptides.

In further aspects of this invention, methods and diagnostic kits are provided for detecting tuberculosis in a patient. The methods comprise contacting dermal cells of a patient with one or more of the above polypeptides and detecting an immune

response on the patient's skin. The diagnostic kits comprise one or more of the above polypeptides in combination with an apparatus sufficient to contact the polypeptide(s) with the dermal cells of a patient.

In yet another aspect, methods are provided for detecting tuberculosis in a patient, such methods comprising contacting dermal cells of a patient with one or more polypeptides encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, the complements of said sequences, or nucleotide sequences that hybridize to a sequence selected from the group consisting of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164, and detecting an immune response on the patient's skin. Diagnostic kits for use in such methods are also provided.

In additional aspects of the invention, methods are provided for inhibiting the development of a *Mycobacterium* infection in a patient. In one embodiment, inhibiting the development of a *Mycobacterium* infection comprises administering to a patient an effective amount of a pharmaceutical composition or a vaccine of the invention. In another embodiment, inhibiting the development of a *Mycobacterium* infection in the patient comprises administering to a patient an effective amount of an antibody of the invention. In a preferred embodiment, the *Mycobacterium* infection is a *M. tuberculosis* infection.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B illustrate the stimulation of proliferation and interferon-γ production, respectively, in T cells derived from a first PPD-positive donor (referred to as D7) by recombinant ORF-2 and synthetic peptides to ORF-2.

Figures 2A and 2B illustrate the stimulation of proliferation and interferon-γ production, respectively, in T cells derived from a second PPD-positive donor (referred to as D160) by recombinant ORF-2 and synthetic peptides to ORF-2.

Figure 3 shows the nucleotide sequence of mTTC#3 (SEQ ID NO:145).

Figure 4 shows the amino acid sequence of mTCC#3 (SEQ ID NO:146).

Figures 5 shows the 5' nucleotide sequence of P1 (SEQ ID NO:149).

Figures 6 shows the nucleotide sequence of P2 (SEQ ID NO:150).

Figure 7 shows the 3' nucleotide sequence of P3 (SEQ ID NO:151).

Figure 8 shows the nucleotide sequence of P4 (SEQ ID NO:152).

Figure 9 shows the nucleotide sequence of P6 (SEQ ID NO:153)

Figure 10 shows the nucleotide sequence of P7 (SEQ ID NO:154)

Figure 11 shows the nucleotide sequence of P8 (SEQ ID NO:155)

Figure 12 shows the nucleotide sequence of P9 (SEQ ID NO:156)

Figure 13 shows the 5' nucleotide sequence of P10 (SEQ ID NO:157)

Figure 14 shows the 5' nucleotide sequence of P11 (SEQ ID NO:158)

Figure 15 shows the 3' nucleotide sequence of P12 (SEQ ID NO:159)

Figure 16 shows the full length nucleotide and amino acid sequence of MO-1 (SEQ ID NO:160 (nucleotide) and SEQ ID NO: 161 (amino acid).

Figure 17 shows the full length nucleotide and amino acid sequence of MO-2 (SEQ ID NO:162 (nucleotide) and SEQ ID NO: 163 (amino acid).

Figure 18 shows the full length nucleotide sequence of TbH4/XP-1 (MTB48) (SEQ ID NO:164).

SEQ ID NO:1 is the cDNA sequence of Tb224

SEQ ID NO:2 is the cDNA sequence of Tb636

SEQ ID NO:3 is the cDNA sequence of Tb424

SEQ ID NO:4 is the cDNA sequence of Tb436

SEQ ID NO:5 is the cDNA sequence of Tb398

SEQ ID NO:6 is the cDNA sequence of Tb508

SEQ ID NO:7 is the cDNA sequence of Tb441

SEQ ID NO:8 is the cDNA sequence of Tb475

SEQ ID NO:9 is the cDNA sequence of Tb488

SEQ ID NO:10 is the cDNA sequence of Tb465

SEQ ID NO:11 is the cDNA sequence of Tb431

SEQ ID NO:12 is the cDNA sequence of Tb472

SEQ ID NO:13 is the predicted amino acid sequence of Tb224

SEQ ID NO:14 is the predicted amino acid sequence of Tb636

SEQ ID NO:15 is the predicted amino acid sequence of Tb431
SEQ ID NO:16 is the amino acid sequence of Tb424 ORF-1
SEQ ID NO:17 is the amino acid sequence of Tb424 ORF-2
SEQ ID NO:18 is the amino acid sequence of Tb436 ORF-1
SEQ ID NO:19 is the amino acid sequence of Tb436 ORF-2
SEQ ID NO:20 is the amino acid sequence of Tb398 ORF-1
SEQ ID NO:21 is the amino acid sequence of Tb398 ORF-2
SEQ ID NO:22 is the amino acid sequence of Tb508 ORF-1
SEQ ID NO:23 is the amino acid sequence of Tb508 ORF-2
SEQ ID NO:24 is the amino acid sequence of Tb441 ORF-1
SEQ ID NO:25 is the amino acid sequence of Tb441 ORF-2
SEQ ID NO:26 is the amino acid sequence of Tb475 ORF-1
SEQ ID NO:27 is the amino acid sequence of Tb475 ORF-2
SEQ ID NO:28 is the amino acid sequence of Tb488 ORF-1
SEQ ID NO:29 is the amino acid sequence of Tb488 ORF-2
SEQ ID NO:30 is the amino acid sequence of Tb465 ORF-1
SEQ ID NO:31 is the amino acid sequence of Tb465 ORF-2
SEQ ID NO:32 is the amino acid sequence of Tb424 ORF-U
SEQ ID NO:33 is the amino acid sequence of Tb436 ORF-U
SEQ ID NO:34 is the amino acid sequence of ORF-1-1
SEQ ID NO:35 is the amino acid sequence of ORF-1-2
SEQ ID NO:36 is the amino acid sequence of ORF-1-3
SEQ ID NO:37 is the amino acid sequence of ORF-1-4
SEQ ID NO:38 is the amino acid sequence of ORF-1-5
SEQ ID NO:39 is the amino acid sequence of ORF-1-6
SEQ ID NO:40 is the amino acid sequence of ORF-1-7
SEQ ID NO:41 is the amino acid sequence of ORF-1-8
SEQ ID NO:42 is the amino acid sequence of ORF-1-9
SEQ ID NO:43 is the amino acid sequence of ORF-1-10
SEQ ID NO:44 is the amino acid sequence of ORF-1-11
SEQ ID NO:45 is the amino acid sequence of ORF-1-12
SEQ ID NO:46 is the amino acid sequence of ORF-1-13

SEQ ID NO:47 is the amino acid sequence of ORF-1-14
SEQ ID NO:48 is the amino acid sequence of ORF-1-15
SEQ ID NO:49 is the amino acid sequence of ORF-1-16
SEQ ID NO:50 is the amino acid sequence of ORF-1-17
SEQ ID NO:51 is the amino acid sequence of ORF-2-1
SEQ ID NO:52 is the amino acid sequence of ORF-2-2
SEQ ID NO:53 is the amino acid sequence of ORF-2-3
SEQ ID NO:54 is the amino acid sequence of ORF-2-4
SEQ ID NO:55 is the amino acid sequence of ORF-2-5
SEQ ID NO:56 is the amino acid sequence of ORF-2-6
SEQ ID NO:57 is the amino acid sequence of ORF-2-7
SEQ ID NO:58 is the amino acid sequence of ORF-2-8
SEQ ID NO:59 is the amino acid sequence of ORF-2-9
SEQ ID NO:60 is the amino acid sequence of ORF-2-10
SEQ ID NO:61 is the amino acid sequence of ORF-2-11
SEQ ID NO:62 is the amino acid sequence of ORF-2-12
SEQ ID NO:63 is the amino acid sequence of ORF-2-13
SEQ ID NO:64 is the amino acid sequence of ORF-2-14
SEQ ID NO:65 is the amino acid sequence of ORF-2-15
SEQ ID NO:66 is the amino acid sequence of ORF-2-16
SEQ ID NO:67 is the amino acid sequence of ORF-2-17
SEQ ID NO:68 is the amino acid sequence of ORF-2-18
SEQ ID NO:69 is the amino acid sequence of ORF-2-19
SEQ ID NO:70 is the amino acid sequence of ORF-2-20
SEQ ID NO:71 is the amino acid sequence of ORF-2-21
SEQ ID NO:72 is the amino acid sequence of ORF-2-22
SEQ ID NO:73 is the amino acid sequence of ORF-2-23
SEQ ID NO:74 is the amino acid sequence of ORF-2-24
SEQ ID NO:75 is the amino acid sequence of ORF-2-25
SEQ ID NO:76 is the amino acid sequence of ORF-2-26
SEQ ID NO:77 is the amino acid sequence of ORF-2-27
SEQ ID NO:78 is the amino acid sequence of ORF-2-28

SEQ ID NO:79 is the amino acid sequence of ORF-2-29

SEQ ID NO:80 is the amino acid sequence of ORF-2-30

SEQ ID NO:81-82 are the amino acid sequence of two overlapping peptides to the open reading frame of Tb224

SEQ ID NO:83 is the full-length cDNA sequence of Tb431 (which contains an ORF encoding Mtb-40)

SEQ ID NO:84 is the amino acid sequence of MSF-1

SEQ ID NO:85 is the amino acid sequence of MSF-2

SEQ ID NO:86 is the amino acid sequence of MSF-3

SEQ ID NO:87 is the amino acid sequence of MSF-4

SEQ ID NO:88 is the amino acid sequence of MSF-5

SEQ ID NO:89 is the amino acid sequence of MSF-6

SEQ ID NO:90 is the amino acid sequence of MSF-7

SEQ ID NO:91 is the amino acid sequence of MSF-8

SEQ ID NO:92 is the amino acid sequence of MSF-9

SEQ ID NO:93 is the amino acid sequence of MSF-10

SEQ ID NO:94 is the amino acid sequence of MSF-11

SEQ ID NO:95 is the amino acid sequence of MSF-12

SEQ ID NO:96 is the amino acid sequence of MSF-13

SEQ ID NO:97 is the amino acid sequence of MSF-14

SEQ ID NO:98 is the amino acid sequence of MSF-15

SEQ ID NO:99 is the amino acid sequence of MSF-16

SEQ ID NO:100 is the amino acid sequence of MSF-17

SEQ ID NO:101 is the amino acid sequence of MSF-18

SEQ ID NO:102 is the cDNA sequence of Tb867

SEQ ID NO:103 is the cDNA sequence of Tb391

SEQ ID NO:104 is the cDNA sequence of Tb470

SEQ ID NO:105 is the cDNA sequence of Tb838

SEQ ID NO:106-107 are the cDNA sequences of Tb962

SEQ ID NO:108 is the full-length cDNA sequence of Tb472

SEQ ID NO:109 is the predicted amino acid sequence of the protein encoded by Tb472 (referred to as MSL)

SEQ ID NO:110 is the amino acid sequence of MSL-1
SEQ ID NO:111 is the amino acid sequence of MSL-2.
SEQ ID NO:112 is the amino acid sequence of MSL-3
SEQ ID NO:113 is the amino acid sequence of MSL-4
SEQ ID NO:114 is the amino acid sequence of MSL-5
SEQ ID NO:115 is the amino acid sequence of MSL-6
SEQ ID NO:116 is the amino acid sequence of MSL-7
SEQ ID NO:117 is the amino acid sequence of MSL-8
SEQ ID NO:118 is the amino acid sequence of MSL-9
SEQ ID NO:119 is the amino acid sequence of MSL-10
SEQ ID NO:120 is the amino acid sequence of MSL-11
SEQ ID NO:121 is the amino acid sequence of MSL-12
SEQ ID NO:122 is the amino acid sequence of MSL-13
SEQ ID NO:123 is the amino acid sequence of MSL-14
SEQ ID NO:124 is the amino acid sequence of MSL-15
SEQ ID NO:125 is the DNA sequence of the full-length open reading frame of Tb470 (which encodes Mtb-40)
SEQ ID NO:126 is the determined amino acid sequence of Mtb-40
SEQ ID NO:127 is the cDNA sequence of Tb366
SEQ ID NO:128 is the cDNA sequence of Tb433
SEQ ID NO:129 is the cDNA sequence of Tb439
SEQ ID NO:130-131 are the cDNA sequences of Tb372
SEQ ID NO:132 is the cDNA sequence of Tb390R5C6
SEQ ID NO:133-134 are the cDNA sequences of Tb390R2C11
SEQ ID NO:135 is the 5' cDNA sequence of Y1-26C1
SEQ ID NO:136 is the 5' cDNA sequence of Y1-86C11
SEQ ID NO:137 is the full-length cDNA sequence of hTcc#1
SEQ ID NO:138 is the predicted amino acid sequence of hTcc#1
SEQ ID NO:139 is the cDNA sequence of mTCC#1
SEQ ID NO:140 is the cDNA sequence of mTCC#2
SEQ ID NO:141 is the predicted amino acid sequence of mTCC#1
SEQ ID NO:142 is the predicted amino acid sequence of mTCC#2

SEQ ID NO:143 is the amino acid sequence of MTb9.8

SEQ ID NO:144 is the amino acid sequence of Tb#470

SEQ ID NO:145 is the full length nucleotide sequence of mTTC#3

SEQ ID NO:146 is the predicted amino acid sequence of mTTC#3

SEQ ID NO:147 and 148 are the sequences of primers used to amplify the full-length coding sequence of mTTC#3

SEQ ID NO:149 is the 5' nucleotide sequence of P1

SEQ ID NO:150 is the nucleotide sequence of P2

SEQ ID NO:151 is the 3' nucleotide sequence of P3

SEQ ID NO:152 is the nucleotide sequence of P4

SEQ ID NO:153 is the nucleotide sequence of P6

SEQ ID NO:154 is the nucleotide sequence of P7

SEQ ID NO:155 is the nucleotide sequence of P8

SEQ ID NO:156 is the nucleotide sequence of P9

SEQ ID NO:157 is the 5' nucleotide sequence of P10

SEQ ID NO:158 is the 5' nucleotide sequence of P11

SEQ ID NO:159 is the 3' nucleotide sequence of P12

SEQ ID NO:160 is the full length nucleotide sequence of MO-1

SEQ ID NO:161 is the full length amino acid sequence of MO-1.

SEQ ID NO:162 is the full length nucleotide sequence of MO-2

SEQ ID NO:163 is the full length amino acid sequence of MO-2

SEQ ID NO:164 is the full length nucleotide sequence of TbH4/XP-1 (MTB48).

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

I. INTRODUCTION

As noted above, the present invention is generally directed to compositions and methods for preventing, treating and diagnosing tuberculosis. In particular, the present invention relates to *Mycobacterium* antigens, optionally from a species such as *M. tuberculosis, M. bovis, M. smegmatis, BCG, M. leprae, M. scrofulaceum, M. avium-intracellulare, M. marinum, M. ulcerans, M. kansasii, M. xenopi, M. szulgai, M. fortuitum,*

or *M. chelonei*. In particular, the invention relates to *Mycobacterium* polypeptides and immunogenic fragments thereof, polynucleotides that encode the polypeptides and immunogenic fragments thereof, and methods of using such compositions in the treatment, prevention and diagnosis of *Mycobacterium* infection. In one embodiment of the invention, the polypeptides of the invention are used to diagnose tuberculosis. In another embodiment of the invention, the polypeptides of the invention are used to induce an immune response in a patient in order to prevent *Mycobacterium* infection, and in particular tuberculosis, or to reduce the probability of pathological responses typical of *Mycobacterium* infection, and in particular tuberculosis, in a patient. In another embodiment of the invention, the polynucleotides of the invention are used to produce DNA vaccines, or for diagnostic purposes.

II. DEFINITIONS

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs). The term also encompasses ribonucleotides including HnRNA molecules, which contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions, as further described below, preferably such that the immunogenicity of the encoded polypeptide is not diminished relative to the native polypeptide. The effect on

the immunogenicity of the encoded polypeptide may generally be assessed as described herein. The term "variants" also encompasses interspecies homologs. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res. 19*:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem. 260*:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes 8*:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds. Thus, for instance, a polypeptide comprising an immunogenic portion of an antigen may consist entirely of the immunogenic portion, or may contain additional sequences. The additional sequences may be derived from the native *Mycobacterium* antigen or may be heterologous, and such sequences may (but need not) be immunogenic.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

The compositions and methods of this invention also encompass variants of the above polypeptides. A polypeptide "variant," as used herein, is a polypeptide that differs from the recited polypeptide only in conservative substitutions and/or modifications, such that the therapeutic and/or immunogenic properties of the polypeptide are retained. Polypeptide variants preferably exhibit at least about 70%, more preferably at least about 90% and most preferably at least about 95% identity to the identified polypeptides. For polypeptides with immunoreactive properties, variants may, alternatively, be identified by modifying the amino acid sequence of one of the above polypeptides, and evaluating the immunoreactivity of the modified polypeptide. For polypeptides useful for the generation of diagnostic binding agents, a variant may be identified by evaluating a modified polypeptide for the ability to generate antibodies that detect the presence or absence of *Mycobacterium* infection, and in particular tuberculosis. Alternatively, variants of the claimed antigens that may be usefully employed in the inventive diagnostic methods may be identified by evaluating modified polypeptides for their ability to detect antibodies present in the sera of *Mycobacterium*-infected patients. Such modified sequences may be prepared and tested using, for example, the representative procedures described herein.

A "conservative substitution" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservative substitutions refers to changes in the nucleic acid sequence that result in nucleic acids encoding identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence

herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservative substitution" where the alteration results in the substitution of an amino acid with a chemically similar amino acid and where the alteration has minimal influence on the immunogenic properties, secondary structure and hydropathic nature of the polypeptide. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

"Immunogenic," as used herein, refers to the ability to elicit an immune response (*e.g.*, cellular or humoral) in a patient, such as a human, and/or in a biological sample (*in vitro*). In particular, antigens that are immunogenic (and immunogenic portions or other variants of such antigens) are recognized by a B-cell and/or a T-cell surface antigen receptor. Antigens that are immunogenic (and immunogenic portions or

other variants of such antigens) are capable of stimulating cell proliferation, interleukin-12 production and/or interferon-γ production in biological samples comprising one or more cells selected from the group of T cells, NK cells, B cells and macrophages, where the cells are derived from an *Mycobacterium*-immune individual. Polypeptides comprising at least an immunogenic portion of one or more *Mycobacterium* antigens may generally be used to detect tuberculosis or to induce protective immunity against tuberculosis in a patient.

"Fusion polypeptide" or "fusion protein" refers to a protein having at least two heterologous polypeptides covalently linked, preferably *Mycobacterium* sp. polypeptides, either directly or via an amino acid linker. The polypeptides forming the fusion protein are typically linked C-terminus to N-terminus, although they can also be linked C-terminus to C-terminus, N-terminus to N-terminus, or N-terminus to C-terminus. The polypeptides of the fusion protein can be in any order. This term also refers to conservatively modified variants, polymorphic variants, alleles, mutants, subsequences, and interspecies homologs of the antigens that make up the fusion protein. *Mycobacterium tuberculosis* antigens are described in Cole *et al.*, *Nature 393*:537 (1998). The complete sequence of the *Mycobacterium tuberculosis* genome can be found at http://www.sanger.ac.uk and at http://www.pasteur.fr/mycdb/ (MycDB).

An adjuvant refers to the components in a vaccine or therapeutic composition that increase the specific immune response to the antigen (*see, e.g.*, Edelman, *AIDS Res. Hum Retroviruses* 8:1409-1411 (1992)). Adjuvants induce immune responses of the Th1-type and Th-2 type response. Th1-type cytokines (e.g., IFN-γ, IL-2, and IL-12) tend to favor the induction of cell-mediated immune response to an administered antigen, while Th-2 type cytokines (e.g., IL-4, IL-5, Il-6, IL-10 and TNF-β) tend to favor the induction of humoral immune responses.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and

will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For stringent hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions include: 50% formamide, 5X SSC and 1% SDS incubated at 42° C or 5X SSC and 1% SDS incubated at 65°C, with a wash in 0.2X SSC and 0.1% SDS at 65°C.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cased, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately" stringent hybridization conditions include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that

are the same (*i.e.*, 55%, 60%, 65%, 70%, 75%, or 80% identity, preferably 85%, 90%, 95%, 96%, 97%, 98%, 99% or higher identity over a specified window region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence. Preferably, the identity exists over a region that is at least about 25 amino acids or nucleotides in length, or more preferably over a region that is 50-100 amino acids or nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, *A model of evolutionary change in proteins – Matrices for detecting distant relationships, In*: Dayhoff (ed.) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358 (1978); Hein, *Unified Approach to Alignment and Phylogenes* pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA (1990); Higgins and Sharp, *CABIOS 5*:151-153 (1989); Myers and Muller,

CABIOS 4:11-17 (1988); Robinson, *Comb. Theor 11*:105 (1971); Santou and Nes, *Mol. Biol. Evol. 4*:406-425 (1987); Sneath and Sokal, *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA (1973); Wilbur and Lipman, *Proc. Natl. Acad. Sci. USA 80*:726-730 (1983).

Alternatively, optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math. 2*:482 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol. 48*:443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci. USA 85*:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. (1995 supplement)).

A preferred example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al., Nuc. Acids Res. 25*:3389-3402 (1977) and Altschul *et al., J. Mol. Biol. 215*:403-410 (1990), respectively. BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for the nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X

from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

"Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, *e.g.*, as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab)'$_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The F(ab)'$_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the F(ab)'$_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (*see, e.g.*, *Fundamental Immunology* (Paul ed., 3d ed. (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (*e.g.*, single chain Fv) or those identified using phage display libraries (*see, e.g.*, McCafferty *et al.*, *Nature 348*:552-554 (1990))

As used herein, an antibody, or antigen-binding fragment thereof, is said to "specifically bind" to a polypeptide of interest if it reacts at a detectable level (within, for example, an ELISA) with the polypeptide of interest, and does not react detectably with unrelated proteins under similar conditions. As used herein, "binding" refers to a noncovalent association between two separate molecules such that a complex is formed. The ability to bind may be evaluated by, for example, determining a binding constant for the formation of the complex. The binding constant is the value obtained when the concentration of the complex is divided by the product of the component concentrations. In general, two compounds are said to "bind," in the context of the present invention, when the binding constant for complex formation exceeds about 10^3 l/mol. The binding constant may be determined using methods well known in the art.

As used herein, a "biological sample" is any antibody-containing sample obtained from a patient. Preferably, the sample is whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid or urine. More preferably, the sample is a blood, serum or plasma sample obtained from a patient or a blood supply.

In the context of the present invention, a "patient" refers to any warm-blooded animal, preferably a human. A patient may be afflicted with a disease, or may be free of detectable disease and/or infection.

III. PREPARATION OF *MYCOBACTERIUM* POLYPEPTIDES AND NUCLEIC ACIDS

In general, *Mycobacterium* antigens and DNA sequences encoding such antigens may be prepared using any of a variety of procedures. Here and throughout the specification, the *Mycobacterium* antigens are preferably *M. tuberculosis* antigens.

A. Polynucleotides of the invention

DNA sequences encoding antigens may be identified, for example, by screening an appropriate *Mycobacterium* genomic or cDNA expression library with sera obtained from patients infected with *Mycobacterium*. Alternatively, sera from mice immunized with *Mycobacterium* antigens can be used. In some embodiments, sera is obtained from mice immunized with blood or urine from syngeneic mice infected with *Mycobacterium*. Such screens may generally be performed using techniques well known to those of ordinary skill in the art, such as those described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY (1989).

DNA sequences encoding the antigens of the present invention may also be obtained by screening an appropriate *Mycobacterium* cDNA or genomic DNA library for DNA sequences that hybridize to degenerate oligonucleotides derived from partial amino acid sequences of isolated antigens. Degenerate oligonucleotide sequences for use in such a screen may be designed and synthesized, and the screen may be performed as described, for example, in Sambrook *et al.*, *supra*, and references cited therein.

For hybridization techniques, a partial sequence may be labeled (*e.g.*, by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe (*see* Sambrook *et al.*, *supra*). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences are then assembled into a single

contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

Amplification techniques may also be employed, using the above oligonucleotides in methods well known in the art, to isolate a nucleic acid probe from a cDNA or genomic library. The library screen for obtaining a full length coding sequence from a partial cDNA sequence may then be performed using the isolated probe. Within such techniques, amplification is generally performed via PCR. Any of a variety of commercially available kits may be used to perform the amplification step. Primers may be designed using, for example, software well known in the art. Primers are preferably 22-30 nucleotides in length, have a GC content of at least 50% and anneal to the target sequence at temperatures of about 68°C to 72°C. The amplified region may be sequenced and overlapping sequences assembled into a contiguous sequence.

One such amplification technique is inverse PCR (*see* Triglia *et al.*, *Nucl. Acids Res. 16*:8186 (1988)), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Optionally, capture PCR (Lagerstrom *et al.*, *PCR Methods Applic. 1*:111-119 (1991)) and walking PCR (Parker *et al.*, *Nucl. Acids. Res. 19*:3055-60 (1991)) can also be used. Methods for amplification further include the ligase chain reaction (LCR; *see, e.g.*, EP patent application publication 320, 308), the Qbeta Replicase method (*see, e.g.*, PCT/US87/00880), the isothermal amplification method, the Strand Displacement Amplification (SDA), the cyclic probe reaction (CPR), the transcription-based amplification systems (TAS; *see, e.g.*, PCT/US88/10315), as well as other methods

known to those of skill in the art (*see, e.g.*, GB patent application No. 2,202,328; PCT/US89/01025; and EP patent application publication No. 329,822). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.*, NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence.

Polynucleotide variants may generally be prepared by any method known in the art, including chemical synthesis by, for example, solid phase phosphoramidite chemical synthesis. Modifications in a polynucleotide sequence may also be introduced using standard mutagenesis techniques, such as oligonucleotide-directed site-specific mutagenesis (*see* Adelman *et al.*, *DNA 2*:183 (1983)). Alternatively, RNA molecules may be generated by *in vitro* or *in vivo* transcription of DNA sequences encoding a *Mycobacterium* polypeptide, or portion thereof, provided that the DNA is incorporated into a vector with a suitable RNA polymerase promoter (such as T7 or SP6). Certain portions may be used to prepare an encoded polypeptide, as described *infra*. In addition, or alternatively, a portion may be administered to a patient such that the encoded polypeptide is generated *in vivo* (*e.g.*, by transfecting antigen-presenting cells, such as dendritic cells, with a cDNA construct encoding a *Mycobacterium* polypeptide, and administering the transfected cells to the patient).

A portion of a sequence complementary to a coding sequence (*i.e.*, an antisense polynucleotide) may also be used as a probe or to modulate gene expression. cDNA constructs that can be transcribed into antisense RNA may also be introduced into cells or tissues to facilitate the production of antisense RNA. An antisense polynucleotide may be used, as described herein, to inhibit expression of a *Mycobacterium* protein. Antisense technology can be used to control gene expression through triple-helix formation, which compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors or regulatory molecules (*see* Gee *et al.*, *In* Huber and Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, NY (1994)). Alternatively, an antisense molecule may be designed to hybridize with a control region of a gene (*e.g.*, promoter, enhancer or transcription initiation site),

and block transcription of the gene; or to block translation by inhibiting binding of a transcript to ribosomes.

A portion of a coding sequence or of a complementary sequence may also be designed as a probe or primer to detect gene expression. Probes may be labeled with a variety of reporter groups, such as radionuclides and enzymes, and are preferably at least 10 nucleotides in length, more preferably at least 20 nucleotides in length and still more preferably at least 30 nucleotides in length. Primers, as noted above, are preferably 22-30 nucleotides in length.

Any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl- methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

Nucleotide sequences as described herein may be joined to a variety of other nucleotide sequences using established recombinant DNA techniques. For example, a polynucleotide may be cloned into any of a variety of cloning vectors, including plasmids, phagemids, lambda phage derivatives and cosmids. Vectors of particular interest include expression vectors, replication vectors, probe generation vectors and sequencing vectors. In general, a vector will contain an origin of replication functional in at least one organism, convenient restriction endonuclease sites and one or more selectable markers. Other elements will depend upon the desired use, and will be apparent to those of ordinary skill in the art.

Within certain embodiments, polynucleotides may be formulated so as to permit entry into a cell of a mammal, and expression therein. Such formulations are particularly useful for therapeutic purposes, as described *infra*. Those of ordinary skill in the art will appreciate that there are many ways to achieve expression of a polynucleotide in a target cell, and any suitable method may be employed. For example, a polynucleotide may be incorporated into a viral vector such as, but not limited to, adenovirus, adeno-associated virus, retrovirus, or vaccinia or other pox virus (*e.g.*, avian pox virus). The polynucleotides may also be administered as naked plasmid vectors. Techniques for incorporating DNA into such vectors are well known to those of ordinary

skill in the art. A retroviral vector may additionally transfer or incorporate a gene for a selectable marker (to aid in the identification or selection of transduced cells) and/or a targeting moiety, such as a gene that encodes a ligand for a receptor on a specific target cell, to render the vector target specific. Targeting may also be accomplished using an antibody, by methods known to those of ordinary skill in the art.

Other formulations for therapeutic purposes include colloidal dispersion systems, such as macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. A preferred colloidal system for use as a delivery vehicle *in vitro* and *in vivo* is a liposome (*i.e.*, an artificial membrane vesicle). The preparation and use of such systems is well known in the art.

B. Polypeptides of the invention

Within the context of the present invention, polypeptides may comprise at least an immunogenic portion of a *Mycobacterium* antigen, or a variant thereof, as described herein. As noted above, a *Mycobacterium* antigen is a protein that is expressed by cells infected with *Mycobacterium*. In a preferred embodiment the *Mycobacterium* antigen is a *Mycobacterium tuberculosis* antigen. Proteins that are *Mycobacterium* antigens also react detectably within an immunoassay (such as an ELISA) with antisera from a patient infected with *Mycobacterium*, and preferably with *M. tuberculosis*. Polypeptides as described herein may be of any length. Additional sequences derived from the native protein and/or heterologous sequences may be present, and such sequences may (but need not) possess further immunogenic or antigenic properties.

Genomic or cDNA libraries derived from *Mycobacterium*, and preferably from *M. tuberculosis*, may be screened directly using peripheral blood mononuclear cells (PBMCs) or T cell lines or clones derived from one or more *Mycobacterium*-immune individuals. In a preferred embodiment, the *Mycobacterium*-immune individuals are *M. tuberculosis*-immune individuals. Direct library screens may generally be performed by assaying pools of expressed recombinant proteins for the ability to induce proliferation and/or interferon-γ production in T cells derived from a *Mycobacterium*-immune individual. Potential T cell antigens may be first selected based on antibody reactivity, as described above. Purified antigens are then evaluated for their ability to elicit an

appropriate immune response (*e.g.*, cellular) using, for example, the representative methods described *infra*. Immunogenic antigens may then be partially sequenced using techniques such as traditional Edman chemistry (*see* Edman and Berg, *Eur. J. Biochem. 80*:116-132 (1967)).

Immunogenic antigens may also be produced recombinantly using a DNA sequence that encodes the antigen, which has been inserted into an appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence, and expressed in an appropriate host. Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in Sambrook *et al.*, *supra*; and Ausubel *et al.*, *supra*.

Polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

Portions and other variants of *Mycobacterium* antigens may be generated by synthetic or recombinant means. Synthetic polypeptides having fewer than about 100 amino acids, and generally fewer than about 50 amino acids, may be generated using techniques well known in the art. For example, such polypeptides may be synthesized using any of the commercially available solid-phase techniques, such as the Merrifield solid-phase synthesis method, where amino acids are sequentially added to a growing amino acid chain (*see* Merrifield, *J. Am. Chem. Soc. 85*:2149-2146 (1963)). Equipment for automated synthesis of polypeptides is commercially available from suppliers such as Perkin Elmer/Applied BioSystems Division, Inc., Foster City, CA, and may be operated according to the manufacturer's instructions. Variants of a native antigen may generally be prepared using standard mutagenesis techniques, such as oligonucleotide-directed site-

specific mutagenesis. Sections of the DNA sequence may also be removed using standard techniques to permit preparation of truncated polypeptides.

Recombinant polypeptides containing portions and/or variants of a native antigen may be readily prepared from a DNA sequence encoding the polypeptide using a variety of techniques well known to those of ordinary skill in the art. For example, supernatants from suitable host/vector systems which secrete recombinant protein into culture media may be first concentrated using a commercially available filter. Following concentration, the concentrate may be applied to a suitable purification matrix such as an affinity matrix or an ion exchange resin. Finally, one or more reverse phase HPLC steps can be employed to further purify a recombinant protein.

Any of a variety of expression vectors known to those of ordinary skill in the art may be employed to express recombinant polypeptides of the present invention. Expression may be achieved in any appropriate host cell (*e.g.*, prokaryotic, yeast and higher eukaryotic cell) that has been transformed or transfected with an expression vector containing a DNA molecule that encodes a recombinant polypeptide. Suitable expression vector/host systems include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus (CaMV); tobacco mosaic virus (TMV)) or with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmids); or animal cell systems. Examples of expression vectors for use in bacterial systems include, *e.g.*, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene) and pIN vectors (*see* Van Heeke and Schuster, *J. Biol. Chem. 264*:5503-5509 (1989)). In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used (*see, e.g.*, Ausubel *et al.*, *supra*; and Grant *et al.*, *Methods Enzymol. 153*:516-544 (1987)). In cases where plant expression vectors are used, the expression of sequences encoding polypeptides may be driven by any of a number of promoters, including, but not limited to, the 35S and 19S promoters of CaMV, the omega leader sequence from TMV (Takamatsu, *EMBO J.* 6:307-311 (1987)), as well as plant promoters such as the small subunit of RUBISCO or heat-shock promoters (Coruzzi *et al.*, *EMBO J. 3*:1671-1680

(1984); Broglie *et al.*, *Science 224*:838-843 (1984); and Winter *et al.*, *Results Probl. Cell Differ. 17*:85-105 (1991)). A variety of expression vectors are also available for expression in insect systems. For example, suitable vectors for expression in *Spodoptera frugiperda* cells or in *Trichoplusia* include, but are not limited to, the *Autographa californica* nuclear polyhedrosis virus (AcNPV). Furthermore, viral-based expression systems can also be used to express the polypeptide(s) of interest in mammalian host cells. Preferably, the host cells employed are *E. coli*, yeast or mammalian cell lines, such as COS or CHO. The DNA sequences expressed in this manner may encode naturally occurring antigens, portions of naturally occurring antigens, or other variants thereof.

In general, regardless of the method of preparation, the polypeptides disclosed herein are prepared in substantially pure form. Preferably, the polypeptides are at least about 80% pure, more preferably at least about 90% pure and most preferably at least about 99% pure. For use in the methods described herein, however, such substantially pure polypeptides may be combined.

In one embodiment, the subject invention discloses polypeptides comprising at least an immunogenic portion of a *M. tuberculosis* antigen (or a variant of such an antigen) that comprises the amino acid sequences encoded by (a) the DNA sequence of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164; (b) the complement of such DNA sequence, or (c) a DNA sequence substantially homologous to the sequence of (a) or (b). In a related embodiment, the present invention provides polypeptides comprising at least an immunogenic portion of a *M. tuberculosis* antigen having the amino acid sequence provided in SEQ ID NO:146, 161, or 163, and variants thereof.

The *Mycobacterium* antigens provided herein include variants that are encoded by DNA sequences which are substantially homologous to one or more of DNA sequences specifically recited herein.

C. Fusion polypeptides

In one embodiment, the present invention provides fusion proteins comprising multiple polypeptides of the invention or, alternatively, a polypeptide of the present invention and a known *Mycobacterium* antigen, preferably a *M. tuberculosis* antigen. Examples of such known *Mycobacterium* antigens include, but are not limited to, e.g., 38 kD antigen described in Andersen and Hansen, *Infect. Immun. 57*:2481-2488

(1989) (Genbank Accession No. M30046) and ESAT-6 previously identified in *M. bovis* (Accession No. U34848) and in *M. tuberculosis* (Sorensen *et al.*, *Infec. Immun. 63*:1710-1717 (1995). Examples of suitable *Mycobacterium* antigens are disclosed in U.S. patent application Nos. 09/056,556, 09/223,040 and 09/287,849, and in U.S. provisional patent application Nos. 60/158,338 and 60/158,425, herein each incorporated by reference. Variants of such fusion proteins are also provided.

The fusion proteins of the present invention may also include a fusion partner which may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or assist in expressing the protein (an expression enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see, e.g.*, Stoute *et al.*, *New Engl. J. Med. 336*:86-91, 1997). Other fusion partners may be selected so as to increase the solubility of the protein or to enable the protein to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the protein.

Within preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza* B (WO 91/18926). Preferably, a protein D derivative comprises approximately the first third of the protein (e.g., the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred embodiments, the first 109 residues of a Lipoprotein D fusion partner are included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells. Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemaglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the LytA gene; *Gene 43*:265-292 (1986)). LYTA

is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (*see Biotechnology 10*:795-798 (1992)). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion protein. A repeat portion is found in the C-terminal region starting at residue 178. A particularly preferred repeat portion incorporates residues 188-305.

The fusion proteins of the present invention may also include a linker peptide between the first and second polypeptides. A peptide linker sequence may be employed to separate, for example, the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion protein using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea *et al.*, *Gene 40*:39-46 (1985); Murphy *et al.*, *Proc. Natl. Acad. Sci. USA 83*:8258-8262 (1986); U.S. Patent Nos. 4,935,233 and 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

Fusion proteins may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion protein is expressed as a recombinant protein in an expression system. Briefly, DNA sequences encoding the polypeptide components may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide

component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion protein that retains the biological activity of both component polypeptides.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptide. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

In general, polypeptides (including fusion proteins) and polynucleotides as described herein are isolated. An "isolated" polypeptide or polynucleotide is one that is removed from its original environment. For example, a naturally-occurring protein is isolated if it is separated from some or all of the coexisting materials in the natural system. Preferably, such polypeptides are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure. A polynucleotide is considered to be isolated if, for example, it is cloned into a vector that is not a part of the natural environment.

D. Immunogenicity of the polypeptides of the invention

Regardless of the method of preparation, the antigens and immunogenic portions thereof described herein have the ability to induce an immunogenic response. More specifically, the antigens have the ability to react with sera obtained from a *Mycobacterium*-infected individual and/or to induce proliferation and/or cytokine production (*i.e.*, interferon-γ and/or interleukin-12 production) in T cells, NK cells, B cells and/or macrophages derived from a *Mycobacterium*-immune individual. Here and throughout the specification, the *Mycobacterium*-immune individual is preferably an *M. tuberculosis*-immune individual.

Reactivity with sera obtained from a *Mycobacterium*-infected individual may be evaluated using, for example, the representative ELISA assays described herein, where an absorbance reading with sera from infected individuals that is at least three standard deviations above the absorbance obtained with sera from uninfected individuals is considered positive.

The selection of cell type for use in evaluating an immunogenic response to a antigen will, of course, depend on the desired response. For example, interleukin-12 production is most readily evaluated using preparations containing B cells and/or macrophages. A *Mycobacterium*-immune individual (*e.g.*, an *M. tuberculosis*-immune individual) is one who is considered to be resistant to the development of the disease (*e.g.*, tuberculosis) by virtue of having mounted an effective T cell response to *Mycobacterium* (*i.e.*, substantially free of disease symptoms). Such individuals may be identified based on a strongly positive (*i.e.*, greater than about 10 mm diameter induration) intradermal skin test response to tuberculosis proteins (PPD) and an absence of any signs or symptoms of, *e.g.*, tuberculosis disease. T cells, NK cells, B cells and macrophages derived from *Mycobacterium*-immune individuals may be prepared using methods known to those of ordinary skill in the art. For example, a preparation of PBMCs (*i.e.*, peripheral blood mononuclear cells) may be employed without further separation of component cells. PBMCs may generally be prepared, for example, using density centrifugation through Ficoll™ (Winthrop Laboratories, NY).

T cells for use in the assays described herein may also be purified directly from PBMCs. Alternatively, an enriched T cell line reactive against mycobacterial proteins, or T cell clones reactive to individual mycobacterial proteins, may be employed. Such T cell clones may be generated by, for example, culturing PBMCs from *Mycobacterium*-immune individuals with mycobacterial proteins for a period of 2-4 weeks. This allows expansion of only the mycobacterial protein-specific T cells, resulting in a line composed solely of such cells. These cells may then be cloned and tested with individual proteins, using methods known to those of ordinary skill in the art, to more accurately define individual T cell specificity. In general, antigens that test positive in assays for proliferation and/or cytokine production (*i.e.*, interferon-γ and/or interleukin- 12 production) performed using T cells, NK cells, B cells and/or macrophages derived from an *Mycobacterium*-immune individual are considered immunogenic. Such assays may be performed, for example, using the representative procedures described *infra*. Immunogenic portions of such antigens may be identified using similar assays, and may be present within the polypeptides described herein.

The ability of a polypeptide (*e.g.*, an immunogenic antigen, or a portion or other variant thereof) to induce cell proliferation is evaluated by contacting the cells (*e.g.*,

T cells and/or NK cells) with the polypeptide and measuring the proliferation of the cells. In general, the amount of polypeptide that is sufficient for evaluation of about 10^5 cells ranges from about 10 ng/ml to about 100 µg/ml and preferably is about 10 µg/ml. The incubation of a polypeptide with cells is typically performed at 37°C for about six days. Following incubation with the polypeptide, the cells are assayed for a proliferative response, which may be evaluated by methods known to those of ordinary skill in the art, such as exposing the cells to a pulse of radiolabeled thymidine and measuring the incorporation of label into cellular DNA. In general, a polypeptide that results in at least a three fold increase in proliferation above background (*i.e.*, the proliferation observed for cells cultured without polypeptide) is considered to be able to induce proliferation.

The ability of a polypeptide to stimulate the production of interferon-γ and/or interleukin-12 in cells may be evaluated by contacting the cells with the polypeptide and measuring the level of interferon-γ or interleukin-12 produced by the cells. In general, the amount of polypeptide that is sufficient for the evaluation of about 10^5 cells ranges from about 10 ng/ml to about 100 µg/ml and preferably is about 10 µg/ml. The polypeptide may, but need not, be immobilized on a solid support, such as a bead or a biodegradable microsphere, such as those described in, *e.g.*, U.S. Patent Nos. 4,897,268 and 5,075,109. The incubation of a polypeptide with the cells is typically performed at 37°C for about six days. Following incubation with the polypeptide, the cells are assayed for interferon-γ and/or interleukin-12 (or one or more subunits thereof) production, which may be evaluated by methods known to those of ordinary skill in the art, such as an enzyme-linked immunosorbent assay (ELISA) or, in the case of the IL-12 P70 heterodimer, a bioassay such as an assay measuring proliferation of T cells. In general, a polypeptide that results in the production of at least 50 pg of interferon-γ per ml of cultured supernatant (containing 10^6-10^5 T cells per ml) is considered able to stimulate the production of interferon-γ. A polypeptide that stimulates the production of at least 10 pg/ml of IL-12 P70 subunit, and/or at least 100 pg/ml of IL-12 P40 subunit, per 10^6 macrophages or B cells (or per 3×10^5 PBMC) is considered able to stimulate the production of IL-12.

In general, immunogenic antigens are those antigens that stimulate proliferation and/or cytokine production (*i.e.*, interferon-γ and/or interleukin-12

production) in T cells, NK cells, B cells and/or macrophages derived from at least about 25% of *Mycobacterium*-immune individuals. Among these immunogenic antigens, polypeptides having superior therapeutic properties may be distinguished based on the magnitude of the responses in the above assays and based on the percentage of individuals for which a response is observed. In addition, antigens having superior therapeutic properties will not stimulate proliferation and/or cytokine production *in vitro* in cells derived from more than about 25% of individuals that are not *Mycobacterium*-immune, thereby eliminating responses that are not specifically due to *Mycobacterium*-responsive cells. Those antigens that induce a response in a high percentage of T cell, NK cell, B cell and/or macrophage preparations from *Mycobacterium*-immune individuals (with a low incidence of responses in cell preparations from other individuals) have superior therapeutic properties.

Antigens with superior therapeutic properties may also be identified based on their ability to diminish the severity of *Mycobacterium* infection in experimental animals, when administered as a vaccine. Suitable vaccine preparations for use on experimental animals are described in detail below. Efficacy may be determined based on the ability of the antigen to provide at least about a 50% reduction in bacterial numbers and/or at least about a 40% decrease in mortality following experimental infection. Suitable experimental animals include, *e.g.*, mice, guinea pigs and primates.

Antigens having superior diagnostic properties may generally be identified based on the ability to elicit a response in an intradermal skin test performed on an individual with active tuberculosis, but not in a test performed on an individual who is not infected with *Mycobacterium*. Skin tests may generally be performed as described below, with a response of at least 5 mm induration considered positive.

Immunogenic portions of *Mycobacterium* antigens may be prepared and identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3d ed., Raven Press, pp. 243-247 (1993) and references cited therein. Such techniques include screening polypeptide portions of the native antigen for immunogenic properties and in particular, *e.g.*, ability to react with antigen-specific antibodies, antisera and/or T-cell lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such

antisera and antibodies may be prepared as described herein, and using well known techniques. The representative ELISAs as well as the proliferation and cytokine production assays described herein may generally be employed in these screens. An immunogenic portion of a polypeptide is a portion that, within such representative assays, generates a signal or an immune response (e.g., proliferation, interferon-γ production and/or interleukin-12 production) that is not substantially less than that generated by the full length polypeptide. In other words, an immunogenic portion of a *Mycobacterium* antigen generates at least about 20%, and preferably about 100%, of the signal and/or immune response induced by the full length antigen in the model ELISA or proliferation assay described herein, respectively. An immunogenic portion may also, or alternatively, stimulate the production of at least about 20%, and preferably about 100%, of the interferon-γ and/or interleukin-12 induced by the full length antigen in the model assay described herein. Such immunogenic portions may also react within such assays at a level that is greater than the reactivity of the full length polypeptide. Such screens may generally be performed using methods well known to those of ordinary skill in the art, such as those described in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988). For use in the methods described herein, substantially pure polypeptides may be combined.

IV. ANTIBODIES

The present invention further provides agents, such as antibodies and antigen-binding fragments thereof, that specifically bind to the polypeptides of the invention. Binding agents may be capable of differentiating between patients infected or not with *Mycobacterium*, and in particular with *M. tuberculosis*, using the representative assays provided *infra*. In other words, antibodies or other binding agents that bind to a *Mycobacterium* antigen will generate a signal indicating the presence of tuberculosis in at least about 20% of patients with the disease, and will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without tuberculosis. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.*, blood, sera, urine, sputum, saliva, *etc.*) from patients with and without tuberculosis (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. It will be apparent that a

statistically significant number of samples with and without the disease should be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will recognize that binding agents may be used in combination to improve sensitivity.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component, an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof.

Antibodies may be prepared by any of a variety of techniques known to those of ordinary skill in the art (*see, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988)). In general, antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the immunogenic polypeptide is initially injected into any of a wide variety of mammals (*e.g.*, mice, rats, rabbits, sheep and goats). In this step, the polypeptides of the invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Polyclonal antibodies raised to a fusion protein of the invention can also be obtained by selecting only those polyclonal antibodies that are specifically immunoreactive with the fusion protein of interest and not with the individual polypeptide components of the fusion protein. This selection may be achieved by subtracting out antibodies that cross-react with the individual polypeptide components of the fusion protein of interest.

Alternatively, antibodies that recognize each or all of the individual polypeptide components of a fusion protein may be useful in the context of the present invention.

Monoclonal antibodies specific for the immunogenic polypeptide of interest may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519 (1976), and improvements thereto. Briefly, these methods involve the preparation of immortal cell lines capable of producing antibodies having the desired specificity (*i.e.*, reactivity with the polypeptide of interest). Such cell lines may be produced, for example, from spleen cells obtained from an animal immunized as described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks, colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell line into the peritoneal cavity of a suitable vertebrate host, such as, *e.g.*, a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

Within certain embodiments, the use of antigen-binding fragments of antibodies may be preferred. Such fragments include Fab fragments, which may be prepared using standard techniques. Briefly, immunoglobulins may be purified from rabbit serum by affinity chromatography on Protein A bead columns (*see* Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988)) and

digested by papain to yield Fab and Fc fragments. The Fab and Fc fragments may be separated by affinity chromatography on protein A bead columns.

Antibodies may be used in diagnostic tests to detect the presence of *Mycobacterium* antigens using assays similar to those detailed *infra* and other techniques well known to those of skill in the art, thereby providing a methods for detecting *Mycobacterium* infection, and in particular tuberculosis, in a patient.

Monoclonal antibodies of the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include, but are not limited to, drugs, toxins, and derivatives thereof. Preferred drugs include, *e.g.*, penicillin, rifampin, isoniazid, pyrazinamide, ethambutol, streptomycin, *etc.* These drugs can be obtained from a natural source or be semisynthetic or synthetic compounds. Preferred toxins include ricin, abrin, *Diphtheria* toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups,

sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, including, *e.g.*, U.S. Patent No. 4,671,958.

Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014), by hydrolysis of derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045), by serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789).

It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways. For example, more than one agent may be coupled directly to an antibody molecule, or linkers that provide multiple sites for attachment can be used. Alternatively, a carrier can be used.

A carrier may bear the agents in a variety of ways, including covalent bonding either directly or via a linker group. Suitable carriers include proteins such as, *e.g.*, albumins (*e.g.*, U.S. Patent No. 4,507,234), peptides and polysaccharides such as, *e.g.*, aminodextran (*e.g.*, U.S. Patent No. 4,699,784). A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (*e.g.*, U.S. Patent Nos. 4,429,008 and 4,873,088).

A variety of routes of administration for the antibodies and immunoconjugates may be used. Typically, administration will be, *e.g.*, intravenous, intramuscular, subcutaneous, intranasal, or buccal. It will be evident that the precise dose of the antibody/immunoconjugate will vary depending upon the antibody used, the antigen density in the cells, and the rate of clearance of the antibody.

V. T CELLS

Immunotherapeutic compositions may also, or alternatively, comprise T cells specific for a *Mycobacterium* antigen. Such cells may generally be prepared *in vitro* or *ex vivo*, using standard procedures. For example, T cells may be isolated from bone marrow, peripheral blood or a fraction of bone marrow or peripheral blood of a patient, using a commercially available cell separation system, such as the CEPRATE™ system, available from CellPro Inc., Bothell WA (see also U.S. Patent Nos. 5,240,856 and 5,215,926; WO 89/06280; WO 91/16116 and WO 92/07243). Alternatively, T cells may be derived from related or unrelated humans, non-human mammals, cell lines or cultures.

T cells may be stimulated with a *Mycobacterium* polypeptide, a polynucleotide encoding a *Mycobacterium* polypeptide and/or an antigen presenting cell (APC) that expresses such a polypeptide. Such stimulation is performed under conditions and for a time sufficient to permit the generation of T cells that are specific for the polypeptide. Preferably, a *Mycobacterium* polypeptide or polynucleotide is present within a delivery vehicle, such as a microsphere, to facilitate the generation of specific T cells.

T cells are considered to be specific for a *Mycobacterium* polypeptide if the T cells kill target cells coated with the polypeptide or expressing a gene encoding the polypeptide. T cell specificity may be evaluated using any of a variety of standard techniques. For example, within a chromium release assay or proliferation assay, a stimulation index of more than two fold increase in lysis and/or proliferation, compared to negative controls, indicates T cell specificity. Such assays may be performed, for example, as described in Chen *et al.*, *Cancer Res.* 54:1065-1070 (1994). Alternatively, detection of the proliferation of T cells may be accomplished by a variety of known techniques. For example, T cell proliferation can be detected by measuring an increased rate of DNA synthesis (*e.g.*, by pulse-labeling cultures of T cells with tritiated thymidine and measuring the amount of tritiated thymidine incorporated into DNA). Contact with a *Mycobacterium* polypeptide (100 ng/ml - 100 μg/ml, preferably 200 ng/ml-25 μg/ml) for 3-7 days should result in at least a two fold increase in proliferation of the T cells. Contact as described above for 2-3 hours should result in activation of the T cells, as measured using standard cytokine assays in which a two fold increase in the level of

cytokine release (*e.g.*, TNF or IFN-γ) is indicative of T cell activation (*see* Coligan *et al.*, Current Protocols in Immunology, vol. 1, Wiley Interscience, Greene (1998)). T cells that have been activated in response to a *Mycobacterium* polypeptide, polynucleotide or polypeptide-expressing APC may be $CD4^+$ and/or $CD8^+$. *Mycobacterium* polypeptide-specific T cells may be expanded using standard techniques. Within preferred embodiments, the T cells are derived from a patient, or from a related or unrelated donor, and are administered to the patient following stimulation and expansion.

For therapeutic purposes, $CD4^+$ or $CD8^+$ T cells that proliferate in response to a *Mycobacterium* polypeptide, polynucleotide or APC can be expanded in number either *in vitro* or *in vivo*. Proliferation of such T cells *in vitro* may be accomplished in a variety of ways. For example, the T cells can be re-exposed to a *Mycobacterium* polypeptide (*e.g.*, a short peptide corresponding to an immunogenic portion of such a polypeptide) with or without the addition of T cell growth factors, such as interleukin-2, and/or stimulator cells that synthesize a *Mycobacterium* polypeptide. Alternatively, one or more T cells that proliferate in the presence of a *Mycobacterium* polypeptide can be expanded in number by cloning. Methods for cloning cells are well known in the art, and include limiting dilution. Following expansion, the cells may be administered back to the patient as described, for example, by Chang *et al.*, *Crit. Rev. Oncol. Hematol.* 22:213 (1996).

VI. DIAGNOSTIC ASSAYS

A. Diagnostic assays with *Mycobacterium* polypeptides

In another aspect, the present invention provides methods for using the polypeptides described above to diagnose *Mycobacterium* infection, and in particular tuberculosis. In this aspect, methods are provided for detecting *Mycobacterium* infection in a biological sample, using one or more of the above polypeptides, alone or in combination. In embodiments in which multiple polypeptides are employed, polypeptides other than those specifically described herein, such as the 38 kD antigen described above, may be included. The polypeptide(s) are used in an assay, as described *infra*, to determine the presence or absence of antibodies to the polypeptide(s) in a biological sample (*e.g.*, whole blood, sputum, serum, plasma, saliva, cerebrospinal fluid, urine, *etc.*) relative to a predetermined cut-off value. The presence of such antibodies

indicates previous sensitization to mycobacterial antigens which may be indicative of *Mycobacterium* infection, and in particular tuberculosis.

In embodiments in which more than one polypeptide is employed, the polypeptides used are preferably complementary (*i.e.*, one component polypeptide will tend to detect infection in samples where the infection would not be detected by another component polypeptide). Complementary polypeptides may generally be identified by using each polypeptide individually to evaluate serum samples obtained from a series of patients known to be infected with *Mycobacterium*. After determining which samples test positive (as described below) with each polypeptide, combinations of two or more polypeptides may be formulated that are capable of detecting infection in most, or all, of the samples tested. Such polypeptides are complementary. For example, approximately 25-30% of sera from tuberculosis-infected individuals are negative for antibodies to any single protein, such as the above-mentioned 38 kD antigen. Complementary polypeptides may, therefore, be used in combination with the 38 kD antigen to improve sensitivity of a diagnostic test.

There are a variety of assay formats known to those of ordinary skill in the art for using one or more polypeptides to detect antibodies in a sample. *See, e.g.*, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988), which is incorporated herein by reference. In general, the presence or absence of tuberculosis in a patient may be determined by (a) contacting a biological sample obtained from a patient with one or more polypeptides or fusion proteins of the invention; (b) detecting in the sample a level of antibody that binds to the polypeptide(s) or the fusion protein(s); and (c) comparing the level of antibody with a predetermined cut-off value.

In a preferred embodiment, the assay involves the use of a polypeptide immobilized on a solid support to bind to and remove the antibody from the sample. The bound antibody may then be detected using a detection reagent that contains a reporter group. Suitable detection reagents include antibodies that bind to the antibody/polypeptide complex and free polypeptide labeled with a reporter group (e.g., in a semi-competitive assay). Alternatively, a competitive assay may be utilized, in which an antibody that binds to the polypeptide of interest is labeled with a reporter group and allowed to bind to the immobilized antigen after incubation of the antigen with the

sample. The extent to which components of the sample inhibit the binding of the labeled antibody to the polypeptide is indicative of the reactivity of the sample with the immobilized polypeptide.

The solid support may be any solid material known to those of ordinary skill in the art to which the antigen may be attached. For example, the solid support may be a test well in a microtiter plate or a nitrocellulose or other suitable membrane. Alternatively, the support may be a bead or disc, such as glass, fiberglass, latex or a plastic material such as polystyrene or polyvinylchloride. The support may also be a magnetic particle or a fiber optic sensor, such as those disclosed, for example, in U.S. Patent No. 5,359,681.

The polypeptides may be bound to the solid support using a variety of techniques known to those of ordinary skill in the art, which are amply described in the patent and scientific literature. In the context of the present invention, the term "bound" refers to both noncovalent association, such as adsorption, and covalent attachment (which may be a direct linkage between the antigen and functional groups on the support or may be a linkage by way of a cross-linking agent). Binding by adsorption to a well in a microtiter plate or to a membrane is preferred. In such cases, adsorption may be achieved by contacting the polypeptide, in a suitable buffer, with the solid support for a suitable amount of time. The contact time varies with temperature, but is typically between about 1 hour and 1 day. In general, contacting a well of a plastic microtiter plate (such as polystyrene or polyvinylchloride) with an amount of polypeptide ranging from about 10 ng to about 1 µg, and preferably about 100 ng, is sufficient to bind an adequate amount of antigen.

Covalent attachment of the polypeptide of interest to a solid support may generally be achieved by first reacting the support with a bifunctional reagent that reacts with both the support and a functional group, such as a hydroxyl or amino group, on the polypeptide. For example, the polypeptide may be bound to supports having an appropriate polymer coating using benzoquinone or by condensation of an aldehyde group on the support with an amine and an active hydrogen on the polypeptide (see, *e.g.*, Pierce Immunotechnology Catalog and Handbook, at A12-A13 (1991)).

In certain embodiments, the assay is an enzyme linked immunosorbent assay (ELISA). This assay may be performed by first contacting a polypeptide antigen

that has been immobilized on a solid support, commonly the well of a microtiter plate, with the sample, such that antibodies present within the sample that recognize the polypeptide of interest are allowed to bind to the immobilized polypeptide. Unbound sample is then removed from the immobilized polypeptide and a detection reagent capable of binding to the immobilized antibody-polypeptide complex is added. The amount of detection reagent that remains bound to the solid support is then determined using a method appropriate for the specific detection reagent.

More specifically, once the polypeptide is immobilized on the support as described above, the remaining protein binding sites on the support are typically blocked. Any suitable blocking agent known to those of ordinary skill in the art, such as bovine serum albumin or TWEEN 20™ (Sigma Chemical Co., St. Louis, MO), may be employed. The immobilized polypeptide is then incubated with the sample, and the antibody is allowed to bind to the antigen. The sample may be diluted with a suitable diluent, such as phosphate-buffered saline (PBS) prior to incubation. In general, an appropriate contact time (*i.e.*, incubation time) is a period of time that is sufficient to detect the presence of antibody within a *Mycobacterium*-infected sample. Preferably, the contact time is sufficient to achieve a level of binding that is at least 95% of that achieved at equilibrium between bound and unbound antibody. Those of ordinary skill in the art will recognize that the time necessary to achieve equilibrium may be readily determined by assaying the level of binding that occurs over a period of time. At room temperature, an incubation time of about 30 minutes is generally sufficient.

Unbound sample may then be removed by washing the solid support with an appropriate buffer, such as PBS containing 0.1% TWEEN 20™. Detection reagent may then be added to the solid support. An appropriate detection reagent is any compound that binds to the immobilized antibody-polypeptide complex and that can be detected by any of a variety of means known to those in the art. Preferably, the detection reagent contains a binding agent (such as, for example, Protein A, Protein G, immunoglobulin, lectin or free antigen) conjugated to a reporter group. Preferred reporter groups include enzymes (such as horseradish peroxidase), substrates, cofactors, inhibitors, dyes, radionuclides, luminescent groups, fluorescent groups and biotin. The conjugation of a binding agent to the reporter group may be achieved using standard methods known to those of ordinary skill in the art. Common binding agents may also be

purchased conjugated to a variety of reporter groups from many commercial sources (*e.g.*, Zymed Laboratories, San Francisco, CA, and Pierce, Rockford, IL).

The detection reagent is then incubated with the immobilized antibody-polypeptide complex for an amount of time sufficient to detect the bound antibody. An appropriate amount of time may generally be determined from the manufacturer's instructions or by assaying the level of binding that occurs over a period of time. Unbound detection reagent is then removed and bound detection reagent is detected using the reporter group. The method employed for detecting the reporter group depends upon the nature of the reporter group. For radioactive groups, scintillation counting or autoradiographic methods are generally appropriate. Spectroscopic methods may be used to detect dyes, luminescent groups and fluorescent groups. Biotin may be detected using avidin, coupled to a different reporter group (commonly a radioactive or fluorescent group or an enzyme). Enzyme reporter groups may generally be detected by the addition of substrate (generally for a specific period of time), followed by spectroscopic or other analysis of the reaction products.

To determine the presence or absence of anti-*Mycobacterium* antibodies in the sample, the signal detected from the reporter group that remains bound to the solid support is generally compared to a signal that corresponds to a predetermined cut-off value. In one preferred embodiment, the cut-off value is the average mean signal obtained when the immobilized antigen is incubated with samples from an uninfected patient. In general, a sample generating a signal that is three standard deviations above the predetermined cut-off value is considered positive for *Mycobacterium* infection. In another embodiment, the cut-off value is determined using a Receiver Operator Curve, according to the method of Sackett *et al.*, *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little Brown and Co., pp. 106-107 (1985). Briefly, in this embodiment, the cut-off value may be determined from a plot of pairs of true positive rates (*i.e.*, sensitivity) and false positive rates (100% specificity) that correspond to each possible cut-off value for the diagnostic test result. The cut-off value on the plot that is the closest to the upper left-hand corner (*i.e.*, the value that encloses the largest area) is the most accurate cut-off value, and a sample generating a signal that is higher than the cut-off value determined by this method may be considered positive. Alternatively, the cut-off value may be shifted to the left along the plot, to minimize the false positive rate,

or to the right, to minimize the false negative rate. In general, a sample generating a signal that is higher than the cut-off value determined by this method is considered positive for tuberculosis.

In a related embodiment, the assay is performed in a rapid flow-through or strip test format, wherein the antigen is immobilized on a membrane, such as, e.g., nitrocellulose. In the flow-through test, antibodies within the sample bind to the immobilized polypeptide as the sample passes through the membrane. A detection reagent (e.g., protein A-colloidal gold) then binds to the antibody-polypeptide complex as the solution containing the detection reagent flows through the membrane. The detection of bound detection reagent may then be performed as described above. In the strip test format, one end of the membrane to which the polypeptide is bound is immersed in a solution containing the sample. The sample migrates along the membrane through a region containing the detection reagent and to the area of immobilized polypeptide. The concentration of the detection reagent at the polypeptide indicates the presence of anti-*Mycobacterium* antibodies in the sample. Typically, the concentration of detection reagent at that site generates a pattern, such as a line, that can be read visually. The absence of such a pattern indicates a negative result. In general, the amount of polypeptide immobilized on the membrane is selected to generate a visually discernible pattern when the biological sample contains a level of antibodies that would be sufficient to generate a positive signal in an ELISA, as discussed *supra*. Preferably, the amount of polypeptide immobilized on the membrane ranges from about 25 ng to about 1 μg, and more preferably from about 50 ng to about 500 ng. Such tests can typically be performed with a very small amount (e.g., one drop) of patient serum or blood.

In another aspect, this invention provides methods for using one or more of the polypeptides described above to diagnose *Mycobacterium* infection, and in particular tuberculosis, using a skin test. As used herein, a "skin test" is any assay performed directly on a patient in which a delayed-type hypersensitivity (DTH) reaction (such as swelling, reddening or dermatitis) is measured following intradermal injection of one or more polypeptides as described above. Such injection may be achieved using any suitable device sufficient to contact the polypeptide or polypeptides with dermal cells of the patient, such as a tuberculin syringe or 1 ml syringe. Preferably, the reaction is measured at least 48 hours after injection, more preferably 48-72 hours.

The DTH reaction is a cell-mediated immune response which is greater in patients that have been exposed previously to the test antigen (*i.e.*, the immunogenic portion of the polypeptide employed, or a variant thereof). The response may be measured visually, using a ruler. In general, a response that is greater than about 0.5 cm in diameter, preferably greater than about 1.0 cm in diameter, is a positive response, indicative of *Mycobacterium* infection, which may or may not be manifested as an active disease.

The polypeptides of this invention are preferably formulated, for use in a skin test, as pharmaceutical compositions containing a polypeptide and a physiologically acceptable carrier, as described *infra*. Such compositions typically contain one or more of the above polypeptides in an amount ranging from about 1μg to about 100 μg, preferably from about 10 μg to about 50 μg in a volume of 0.1 ml. Preferably, the carrier employed in such pharmaceutical compositions is a saline solution with appropriate preservatives, such as phenol and/or TWEEN 80™.

In a preferred embodiment, a polypeptide employed in a skin test is of sufficient size such that it remains at the site of injection for the duration of the reaction period. In general, a polypeptide that is at least 9 amino acids in length is sufficient. The polypeptide is also preferably broken down by macrophages within hours of injection to allow presentation to T-cells. Such polypeptides may contain repeats of one or more of the above sequences and/or other immunogenic or non-immunogenic sequences.

Of course, numerous other assay protocols exist that are suitable for use with the polypeptides of the present invention. The above descriptions are intended to be exemplary only.

B. Diagnostic assays with polynucleotides encoding *Mycobacterium* polypeptides

Antibodies may be used in diagnostic tests to detect the presence of *Mycobacterium* antigens using assays similar to those detailed above and other techniques well known to those of skill in the art, thereby providing a method for detecting *Mycobacterium* infection, and in particular tuberculosis, in a patient.

Diagnostic reagents of the present invention may also comprise DNA sequences encoding one or more of the above polypeptides, or one or more portions thereof. Alternatively, *Mycobacterium* infection can be detected based on the level of

mRNA encoding a *Mycobacterium* antigen in a biological sample. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify *Mycobacterium*-specific cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (*i.e.*, hybridizes to) a DNA molecule encoding a polypeptide of the present invention. The presence of the amplified cDNA is then detected using techniques well known in the art, such as gel electrophoresis. Similarly, oligonucleotide probes specific for a DNA molecule encoding a polypeptide of the present invention may be used in a hybridization assay to detect the presence of a polypeptide of the invention in a biological sample.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a *Mycobacterium* antigen that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers and/or probes hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above. Oligonucleotide primers and/or probes which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having the sequence of SEQ ID NO:145, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, and 164. Primers or probes may thus be used to detect *Mycobacterium*-specific sequences in biological samples. DNA probes or primers comprising oligonucleotide sequences described above may be used alone, in combination with each other, or with previously identified sequences, such as the 38 kD antigen discussed above.

Techniques for both PCR based assays and hybridization assays are well known in the art (*see*, for example, Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, *51*:263 (1987); Erlich ed., *PCR Technology*, Stockton Press, NY (1989)).

One preferred assay employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one specific primer generates a cDNA molecule, which may be separated and

visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a test patient and from an individual who is not afflicted with *Mycobacterium* infection. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-infected sample is typically considered positive.

C. Diagnostic assays using the detection of T cells

A *Mycobacterium* infection may also, or alternatively, be detected based on the presence of T cells that specifically react with a *Mycobacterium* protein in a biological sample. Within certain methods, a biological sample comprising $CD4^+$ and/or $CD8^+$ T cells isolated from a patient is incubated with a *Mycobacterium* polypeptide, a polynucleotide encoding such a polypeptide and/or an APC that expresses at least an immunogenic portion of such a polypeptide, and the presence or absence of specific activation of the T cells is detected. Suitable biological samples include, but are not limited to, isolated T cells. For example, T cells may be isolated from a patient by routine techniques (such as by Ficoll/Hypaque density gradient centrifugation of peripheral blood lymphocytes). T cells may be incubated *in vitro* for 2-9 days (typically 4 days) at 37°C with a *Mycobacterium* polypeptide of the invention (at a concentration of, e.g., 5-25 μg/ml). It may be desirable to incubate another aliquot of a T cell sample in the absence of the *Mycobacterium* polypeptide to serve as a control. For $CD4^+$ T cells, activation is preferably detected by evaluating proliferation of the T cells. For $CD8^+$ T cells, activation is preferably detected by evaluating cytolytic activity. A level of proliferation that is at least two fold greater and/or a level of cytolytic activity that is at least 20% greater than in disease-free patients indicates the presence of a *Mycobacterium* infection in the patient.

D. Diagnostic assays for monitoring the progression of the infection

In another embodiment, *Mycobacterium* proteins and polynucleotides encoding such proteins may be used as markers for monitoring the progression of a *Mycobacterium* infection. In this embodiment, assays as described above for the diagnosis of a *Mycobacterium* infection may be performed over time, and the change in the level of reactive polypeptide(s) evaluated. For example, the assays may be performed

every 24-72 hours for a period of 1 month to 6-12 months, and thereafter performed as needed. In general, the *Mycobacterium* infection is progressing in those patients in whom the level of polypeptide detected by the binding agent increases over time. In contrast, the *Mycobacterium* infection is not progressing when the level of reactive polypeptide either remains constant or decreases with time.

As noted above, to improve sensitivity, multiple *Mycobacterium* markers may be assayed within a given sample. It will be apparent that binding agents specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of *Mycobacterium* protein markers may be based on routine experiments to determine combinations that result in optimal sensitivity.

VII. THERAPEUTIC APPLICATIONS

In another aspect, the present invention provides methods for using one or more of the above polypeptides or fusion proteins (or DNA molecules encoding such polypeptides) to induce protective immunity against *Mycobacterium* infection in a patient to either prevent or treat *Mycobacterium* infection, and in particular tuberculosis.

A. Pharmaceutical compositions

In additional embodiments, the present invention concerns formulation of the polypeptides, fusion proteins or DNA molecules disclosed herein in pharmaceutically-acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other modalities of therapy. The pharmaceutical compositions of the invention may comprise one or more polypeptides, each of which may contain one or more of the above sequences (or variants thereof), and a physiologically acceptable carrier.

It will also be understood that, if desired, the polypeptide, fusion protein and nucleic acid molecule compositions disclosed herein may be administered in combination with other agents as well, such as, *e.g.*, other proteins or polypeptides or various pharmaceutically-active agents. In particular, such pharmaceutical compositions may also contain other *Mycobacterium* antigens, either incorporated into a combination polypeptide or present within a separate polypeptide. In fact, there is virtually no limit to other components that may also be included, given that the additional agents do not cause

a significant adverse effect upon contact with the target cells or host tissues. The compositions may thus be delivered along with various other agents as required in the particular instance. Such compositions may be purified from host cells or other biological sources, or alternatively may be chemically synthesized as described herein. Likewise, such compositions may further comprise substituted or derivatized RNA or DNA compositions.

Formulation of pharmaceutically-acceptable excipients and carrier solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including *e.g.*, oral, parenteral, intravenous, intranasal, and intramuscular administration and formulation.

1. Oral administration

In certain applications, the pharmaceutical compositions disclosed herein may be delivered *via* oral administration to an animal. As such, these compositions may be formulated with an inert diluent or with an assimilable edible carrier, or they may be enclosed in hard- or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet.

The active compounds may even be incorporated with excipients and used in the form of ingestible tablets, buccal tables, troches, capsules, elixirs, suspensions, syrups, wafers, and the like (*see, e.g.*, Mathiowitz *et al.*, *Nature 386*:410-414 (1997); Hwang *et al.*, *Crit Rev Ther Drug Carrier Syst. 15*:243-84 (1998); U.S. Patent Nos. 5,641,515; 5,580,579; and 5,792,451). The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. A syrup of elixir may contain the active compound sucrose